

A4_ComponentesPrincipales2

Josue Salvador Cano Martinez

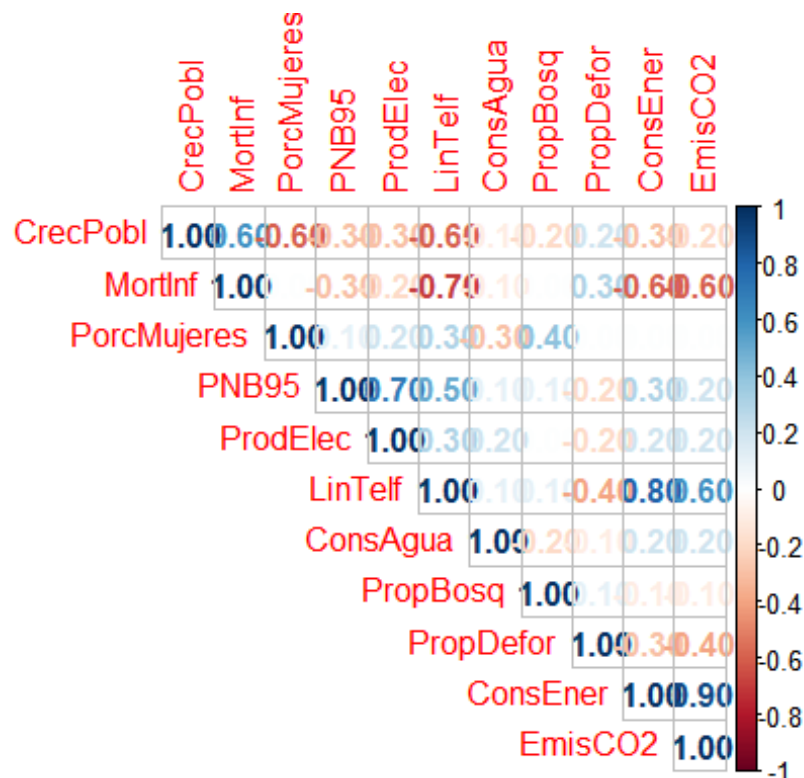
2022-10-18

```
# Cargar datos y generar matrices
M=read.csv("países_mundo.csv")
#M=read.csv("C:/Users/josue/Documents/OneDrive - Instituto Tecnológico y
de Estudios Superiores de
Monterrey/InteligenciaArtificialCienciaDatos/Estadística/países_mundo.csv
")

# Matriz de correlaciones
library(corrplot)

## corrplot 0.92 loaded

correlacion<-round(cor(M), 1)
corrplot(correlacion, method="number", type="upper")
```



```
# Matriz de varianza-covarianza
mcov = cov(M)
print(mcov)
```

```
##          CrecPobl      MortInf      PorcMujeres      PNB95
## CrecPobl      1.538298e+00  2.195026e+01 -6.078026e+00 -8.933379e+04
## MortInf      2.195026e+01  1.032859e+03 -9.249342e+00 -2.269332e+06
## PorcMujeres -6.078026e+00 -9.249342e+00  7.698322e+01  2.813114e+05
## PNB95      -8.933379e+04 -2.269332e+06  2.813114e+05  4.999786e+10
## ProdElec     -4.973964e+04 -1.043435e+06  2.260248e+05  2.247791e+10
## LinTelf     -1.369079e+02 -4.381366e+03  4.499750e+02  2.039550e+07
## ConsAgua     -4.827092e+01 -1.288211e+03 -1.568313e+03  1.097481e+07
## PropBosq     -3.887018e+00 -1.466316e+01  6.517895e+01  2.474311e+05
## PropDefor     3.361974e-01  1.276296e+01  2.680592e-01 -5.806203e+04
## ConsEner     -8.384169e+02 -4.442568e+04  2.855207e+02  1.415628e+08
## EmisCO2     -1.137877e+00 -9.485500e+01 -2.150132e+00  2.501673e+05
##          ProdElec      LinTelf      ConsAgua      PropBosq
## CrecPobl     -4.973964e+04 -1.369079e+02 -4.827092e+01  -3.887018
## MortInf     -1.043435e+06 -4.381366e+03 -1.288211e+03  -14.663158
## PorcMujeres  2.260248e+05  4.499750e+02 -1.568313e+03   65.178947
## PNB95      2.247791e+10  2.039550e+07  1.097481e+07 247431.122807
## ProdElec     1.821909e+10  7.583050e+06  1.399817e+07  70359.785965
## LinTelf      7.583050e+06  3.841247e+04  1.193110e+04   248.715789
## ConsAgua     1.399817e+07  1.193110e+04  3.301981e+05 -2220.757895
## PropBosq     7.035979e+04  2.487158e+02 -2.220758e+03   401.003509
## PropDefor    -3.180340e+04 -9.940461e+01 -6.743793e+01    2.625263
## ConsEner     6.801296e+07  3.426262e+05  2.092242e+05 -5153.438596
## EmisCO2     1.392779e+05  6.385700e+02  4.869328e+02  -12.897193
##          PropDefor      ConsEner      EmisCO2
## CrecPobl     3.361974e-01 -8.384169e+02  -1.137877
## MortInf     1.276296e+01 -4.442568e+04  -94.855000
## PorcMujeres  2.680592e-01  2.855207e+02  -2.150132
## PNB95      -5.806203e+04  1.415628e+08 250167.323509
## ProdElec    -3.180340e+04  6.801296e+07 139277.888640
## LinTelf     -9.940461e+01  3.426262e+05  638.570000
## ConsAgua    -6.743793e+01  2.092242e+05  486.932763
## PropBosq     2.625263e+00 -5.153439e+03  -12.897193
## PropDefor    1.817253e+00 -1.051522e+03  -2.632487
## ConsEner    -1.051522e+03  5.014395e+06 10286.159781
## EmisCO2     -2.632487e+00  1.028616e+04  27.268614
```

valores y vectores propios de matriz de covarianza

```
val = eigen(mcov)$values
```

```
cat("Valores matriz covarianza \n", val, "\n")
```

```
## Valores matriz covarianza
```

```
## 61635762916 6581612268 4636256 310723.2 12160.15 513.7767 362.7885
45.42082 5.800868 1.43802 0.4768083
```

```
vec =eigen(mcov)$vectors
```

```
cat("Vectores matriz de covarianza \n", vec, "\n")
```

```
## Vectores matriz de covarianza
```

```
## -1.658168e-06 -4.048139e-05 5.739096e-06 0.8880376 0.4597636
0.0003504341 0.0002625508 4.089564e-06 -1.073825e-06 0.002547156
```

```

4.643724e-06 4.706785e-07 -1.774254e-05 -1.084543e-05 0.4597632 -
0.8880405 0.0004016179 -0.001122118 7.790843e-06 2.350808e-07
0.0007126782 -1.315731e-06 0.0001263736 0.008225382 0.0001318149
0.002602207 0.0005694896 -0.06194249 -0.04014532 0.001271992 0.0001916177
-0.9972315 -0.002067905 -1.928408e-05 -0.002493257 0.005538307 -
0.0003893588 0.001096305 0.007641174 -0.9991411 0.006435797 4.043796e-05
0.03973568 -5.626049e-05 -0.005537397 -0.09440302 0.03140364 -
0.0003327409 0.0002207819 0.9921404 0.005779514 0.04193316 -0.001809075 -
0.06257295 -0.004236712 0.01243456 0.9917515 0.08552992 -8.621005e-06
1.955408e-05 0.09109622 -0.001087229 0.01721948 0.001758667 0.002639673 -
0.01877994 0.005359089 0.0225802 -0.1136481 -7.566477e-06 1.544658e-05
0.04748682 -0.006863294 -0.9920538 -0.007455427 -0.003764707 -0.001709137
-0.0839081 -0.07891128 0.9856498 1.217248e-05 -2.558998e-05 -0.03416812
0.004698731 -0.1169638 0.01811443 0.001267052 -0.005204823 -0.06778358 -
0.01637836 -0.01468464 -3.971469e-07 1.059471e-06 -0.005379549 7.965261e-
05 0.001416566 0.1283039 0.002262931 -0.9891529 -0.1158091 0.0004264872
0.008241465 4.274451e-07 -1.353881e-06 -0.003409423 3.621425e-05
0.005891758 -0.9859317 0.0002672618 -0.1200519 0.9872887 -0.02092491
0.08344324 2.723996e-07 -2.086857e-07 0.0004944397 0.0004780416 -
0.003748976 -0.1052934 5.906241e-05 -0.08221371

```

valores y vectores propios de matriz de correlaciones

```

val = eigen(correlacion)$values
cat("Valores matriz correlaciones \n", val, "\n")

```

```
## Valores matriz correlaciones
```

```
## 4.081056 1.95246 1.365707 0.8757721 0.8081536 0.7019975 0.6102795
0.3592013 0.1549432 0.09657635 -0.00614596
```

```
vec =eigen(correlacion)$vectors
```

```
cat("Vectores matriz correlaciones \n", vec)
```

```
## Vectores matriz correlaciones
```

```
## -0.3248487 -0.3897603 0.1246114 0.2872485 0.2479505 0.4500074
0.1086279 0.02074124 -0.2437498 0.408602 0.3767421 0.3534153 -0.06516006
-0.5870387 -0.1339892 -0.1318413 -0.05761895 0.3290869 -0.4782568 -
0.1405376 0.2348906 0.2750091 -0.06881855 -0.197846 0.1351162 -0.5422249
-0.6351302 0.1265305 -0.3292559 0.131586 0.004298881 0.2134653 0.2305345
-0.4369213 -0.1532636 0.05297135 -0.2468714 -0.1233358 -0.01886494
0.7161199 0.002791153 0.4146605 -0.0605172 -0.1336032 -0.3082191
0.03055056 0.2071147 -0.2431411 -0.02480592 -0.05804736 0.0748777 -
0.3887821 -0.7229163 -0.2920581 -0.1827693 -0.1842152 -0.1844036
0.1054803 0.04947098 0.09308133 0.02875729 -0.398627 -0.734353 0.4477569
0.02397566 -0.0987753 -0.1493131 -0.578945 -0.5323909 0.2156127 -
0.1913932 0.04434042 -0.1714491 0.1746908 -0.09048299 -0.2833008 -
0.3544625 0.0347437 0.3823753 0.031021 0.3745506 -0.5027289 0.5167871
0.1155083 -0.1408124 -0.02268613 0.1555034 -0.3655955 -0.1533244 -
0.02024823 0.1652217 0.5389878 -0.4467139 -0.4987055 0.0579777 -
0.07291848 -0.0118632 -0.1565469 0.4203437 0.5918942 -0.4136951 0.4176514
0.02711653 -0.05538965 0.2702522 0.2017685 -0.08589631 0.04066383 -
0.4167204 0.06202585 -0.2182931 0.3054971 -0.2804269 -0.0706875

```

```
0.03560554 0.4275893 -0.08570612 -0.007166132 0.1236251 -0.5832365
0.4779524
```

```
# Proporción de varianza explicada por cada componente de matriz de covarianza
```

```
varianzaTotal<-sum(diag(mcov))
lambdas<-eigen(mcov)[1]
lambdas2 <- data.frame(lambdas)
```

```
lambdas2/varianzaTotal
```

```
##          values
## 1  9.034543e-01
## 2  9.647298e-02
## 3  6.795804e-05
## 4  4.554567e-06
## 5  1.782429e-07
## 6  7.530917e-09
## 7  5.317738e-09
## 8  6.657763e-10
## 9  8.502887e-11
## 10 2.107843e-11
## 11 6.989035e-12
```

```
# Acumulación de resultados, matriz de covarianza
```

```
cumsum(mcov)
```

```
## [1] 1.538298e+00 2.348856e+01 1.741054e+01 -8.931638e+04 -
1.390560e+05
## [6] -1.391929e+05 -1.392412e+05 -1.392451e+05 -1.392447e+05 -
1.400832e+05
## [11] -1.400843e+05 -1.400623e+05 -1.390295e+05 -1.390387e+05 -
2.408371e+06
## [16] -3.451806e+06 -3.456188e+06 -3.457476e+06 -3.457491e+06 -
3.457478e+06
## [21] -3.501904e+06 -3.501998e+06 -3.502005e+06 -3.502014e+06 -
3.501937e+06
## [26] -3.220625e+06 -2.994601e+06 -2.994151e+06 -2.995719e+06 -
2.995654e+06
## [31] -2.995653e+06 -2.995368e+06 -2.995370e+06 -3.084704e+06 -
5.354036e+06
## [36] -5.072725e+06 4.999279e+10 7.247070e+10 7.249109e+10
7.250207e+10
## [41] 7.250231e+10 7.250226e+10 7.264382e+10 7.264407e+10
7.264402e+10
## [46] 7.264298e+10 7.264320e+10 9.512111e+10 1.133402e+11
1.133478e+11
## [51] 1.133618e+11 1.133618e+11 1.133618e+11 1.134298e+11
1.134300e+11
## [56] 1.134300e+11 1.134300e+11 1.134300e+11 1.134504e+11
1.134579e+11
```

```
## [61] 1.134580e+11 1.134580e+11 1.134580e+11 1.134580e+11
1.134583e+11
## [66] 1.134583e+11 1.134583e+11 1.134583e+11 1.134583e+11
1.134693e+11
## [71] 1.134833e+11 1.134833e+11 1.134836e+11 1.134836e+11
1.134836e+11
## [76] 1.134839e+11 1.134839e+11 1.134839e+11 1.134839e+11
1.134839e+11
## [81] 1.134841e+11 1.134842e+11 1.134842e+11 1.134842e+11
1.134842e+11
## [86] 1.134842e+11 1.134842e+11 1.134842e+11 1.134842e+11
1.134842e+11
## [91] 1.134842e+11 1.134841e+11 1.134841e+11 1.134841e+11
1.134841e+11
## [96] 1.134841e+11 1.134841e+11 1.134841e+11 1.134841e+11
1.134841e+11
## [101] 1.134840e+11 1.134840e+11 1.136256e+11 1.136936e+11
1.136939e+11
## [106] 1.136942e+11 1.136942e+11 1.136942e+11 1.136992e+11
1.136992e+11
## [111] 1.136992e+11 1.136992e+11 1.136992e+11 1.136994e+11
1.136996e+11
## [116] 1.136996e+11 1.136996e+11 1.136996e+11 1.136996e+11
1.136996e+11
## [121] 1.136996e+11
```

Proporción de varianza explicada por cada componente de matriz de correlación

```
varianzaTotal<-sum(diag(correlacion))
lambdas<-eigen(correlacion)[1]
lambdas2 <- data.frame(lambdas)
```

```
lambdas2/varianzaTotal
```

```
##          values
## 1  0.3710050854
## 2  0.1774963543
## 3  0.1241551414
## 4  0.0796156481
## 5  0.0734685129
## 6  0.0638179509
## 7  0.0554799556
## 8  0.0326546631
## 9  0.0140857438
## 10 0.0087796680
## 11 -0.0005587236
```

Acumulación de resultados, matriz de correlación
cumsum(correlacion)

```
## [1] 1.0 1.6 1.0 0.7 0.4 -0.2 -0.3 -0.5 -0.3 -0.6 -0.8 -0.2 0.8
0.8 0.5
## [16] 0.3 -0.4 -0.5 -0.5 -0.2 -0.8 -1.4 -2.0 -2.0 -1.0 -0.9 -0.7 -0.4
-0.7 -0.3
## [31] -0.3 -0.3 -0.3 -0.6 -0.9 -0.8 0.2 0.9 1.4 1.5 1.6 1.4 1.7
1.9 1.6
## [46] 1.4 1.6 2.3 3.3 3.6 3.8 3.8 3.6 3.8 4.0 3.4 2.7 3.0
3.5 3.8
## [61] 4.8 4.9 5.0 4.6 5.4 6.0 5.9 5.8 5.5 5.6 5.8 5.9 6.9
6.7 6.6
## [76] 6.8 7.0 6.8 6.8 7.2 7.3 7.3 7.4 7.2 8.2 8.3 8.2 8.1
8.3 8.6
## [91] 8.6 8.4 8.2 7.8 7.7 7.8 8.8 8.5 8.1 7.8 7.2 7.2 7.5
7.7 8.5
## [106] 8.7 8.6 8.3 9.3 10.2 10.0 9.4 9.4 9.6 9.8 10.4 10.6 10.5
10.1 11.0
## [121] 12.0
```

¿Qué componentes son los más importantes?

Tomando en consideración los resultados de la proporción de varianza explicada por cada componente de la matriz de correlación, los componentes más importantes son: el primero y el segundo, ya que juntos logran explicar el 54.8% de la varianza.

¿Qué variables son las que más contribuyen a la primera y segunda componentes principales?

Para el componente uno: PorcMujeres, propBosq, propDefor, mortInf; queda la posibilidad de: PNB, prodElec, y crecPob. Por la descripción de las variables podemos llamar a este componente 'índice de riqueza'

Para el componente dos: consEner, emisCO2, consAgua y crecPob. Por la descripción de las variables podemos llamar a este componente 'índice de ruralidad'

¿Por qué lo dice?

Tomando en cuenta la gráfica de agrupación de correlaciones de las variables tomando en cuenta los dos primeros componentes (mostrada en la última sección)

¿Influyen las unidades de las variables?

Sí, las unidades de medida de las variables son muy distintas al relacionarse con diferentes tipos de medidas (porcentajes, kWh, etc), esto ocasionará que la matriz de covarianza y correlación resulten diferentes

Compare los resultados de correlación y de varianza-covarianza, ¿qué concluye?

Las proporción de varianza explicada en la matriz de covarianza resultan ser negativas y esto dará como resultado componentes principales que dificultarán su interpretación. Debido a que los datos que se tienen no están estandarizados, el utilizar la matriz de correlaciones es más correcto porque se le da importancia a todas las variables originales, mismo que equivale a estandarizar cada variables a media cero y varianza uno.

Gráficas respectivas a matriz de varianzas-covarianzas

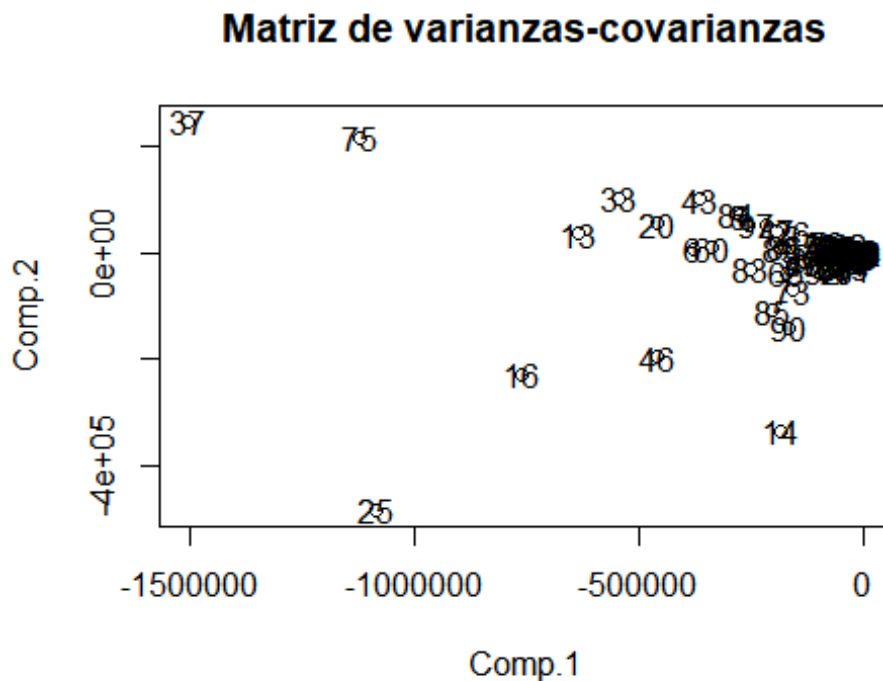
```
library(stats)
library(factoextra)

## Loading required package: ggplot2

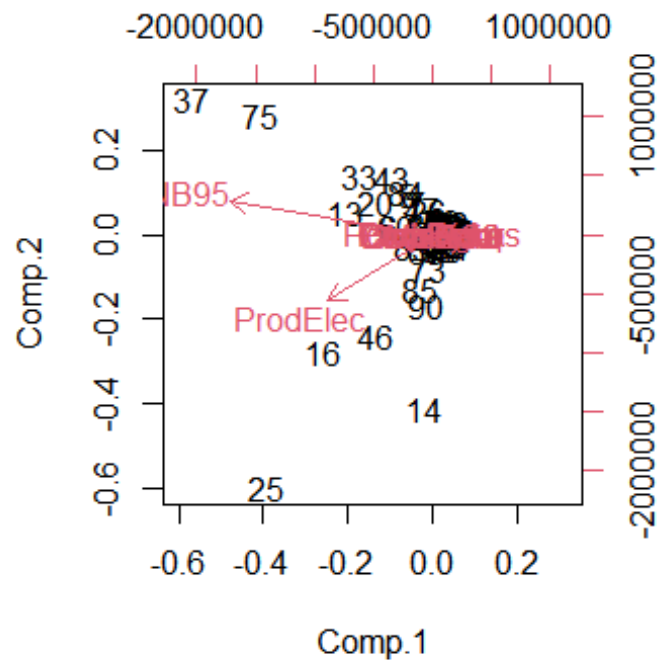
## Warning in register(): Can't find generic `scale_type` in package
## ggplot2 to
## register S3 method.

## Welcome! Want to learn more? See two factoextra-related books at
## https://goo.gl/ve3WBa

library(ggplot2)
datos=M
cpS=princomp(datos,cor=FALSE)
cpaS=as.matrix(datos)%*%cpS$loadings
plot(cpaS[,1:2],type="p", main = "Matriz de varianzas-covarianzas")
text(cpaS[,1],cpaS[,2],1:nrow(cpaS))
```



[illegible]

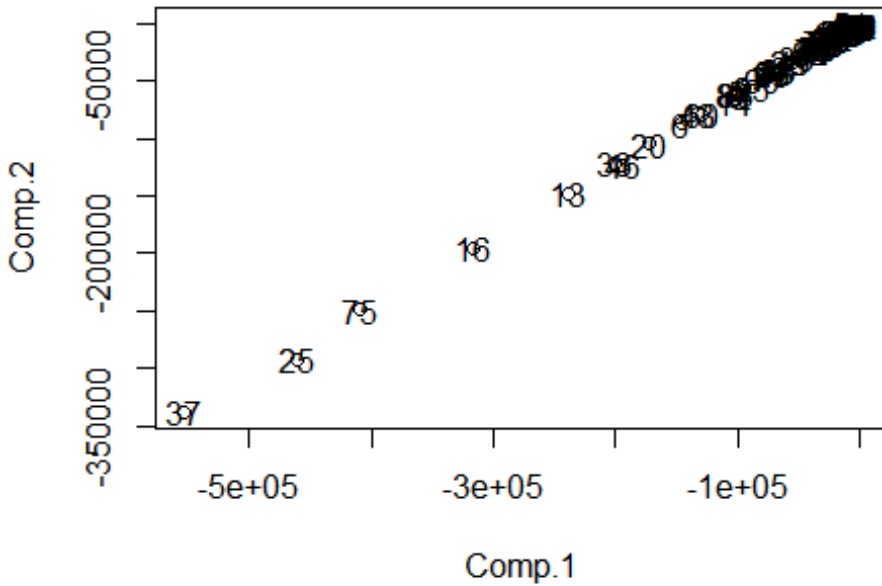


*# Interpretación de resultados en término de agrupación de variables:
Se puede notar que el uso de la matriz de varianzas-covarianzas no resulta útil, por lo que sólo se detallarán las de correlación.*

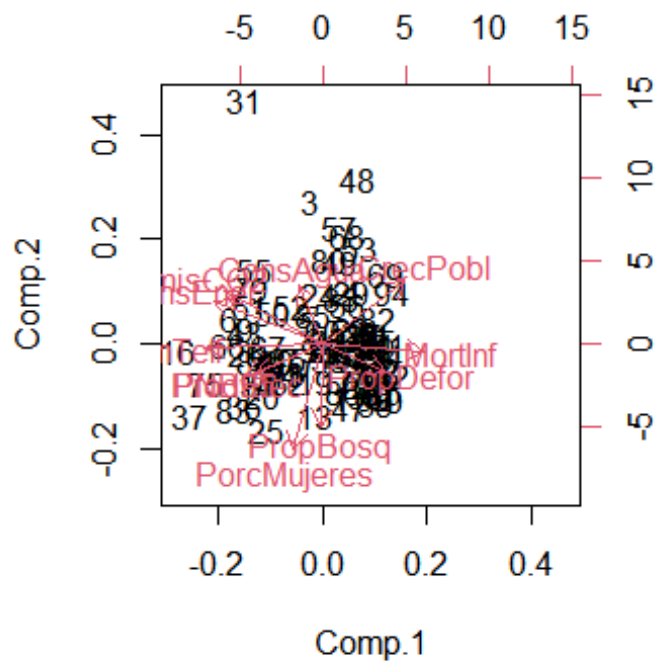
Gráficas respectivas a matriz de correlaciones

```
library(stats)
library(factoextra)
library(ggplot2)
datos=M
cpS=princomp(datos,cor=TRUE)
cpaS=as.matrix(datos)%*%cpS$loadings
plot(cpaS[,1:2],type="p", main = "Matriz de correlaciones")
text(cpaS[,1],cpaS[,2],1:nrow(cpaS))
```

Matriz de correlaciones



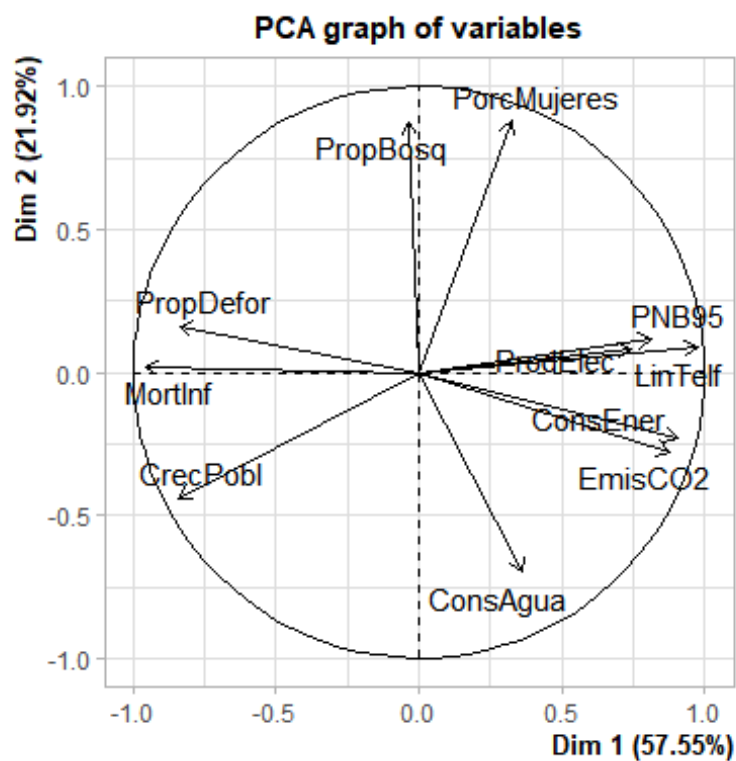
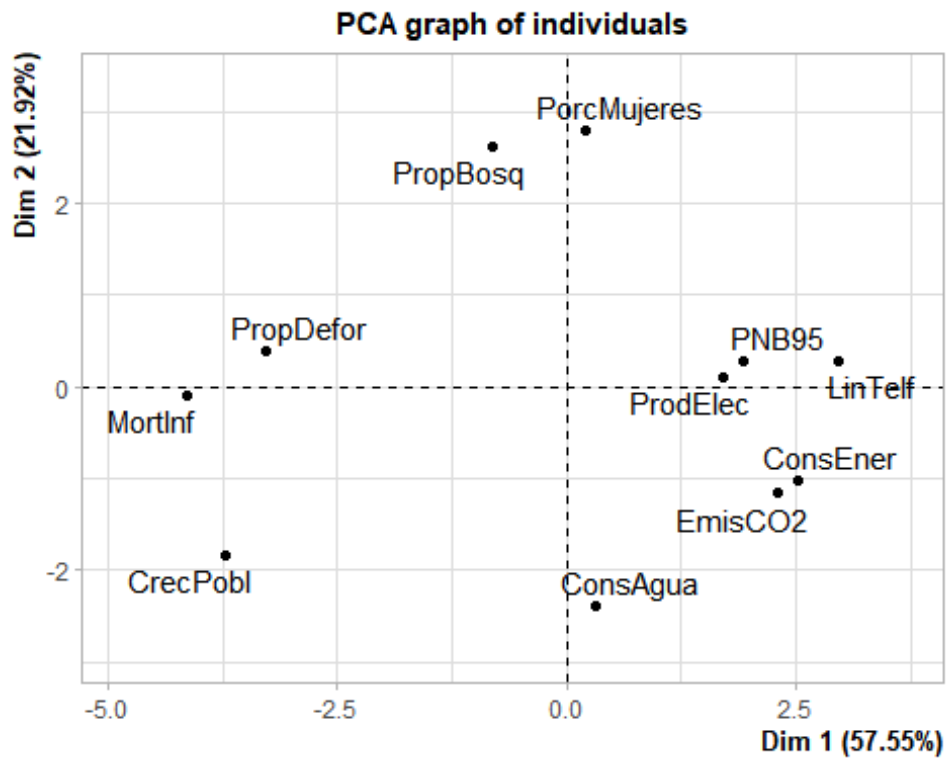
```
biplot(cpS)
```



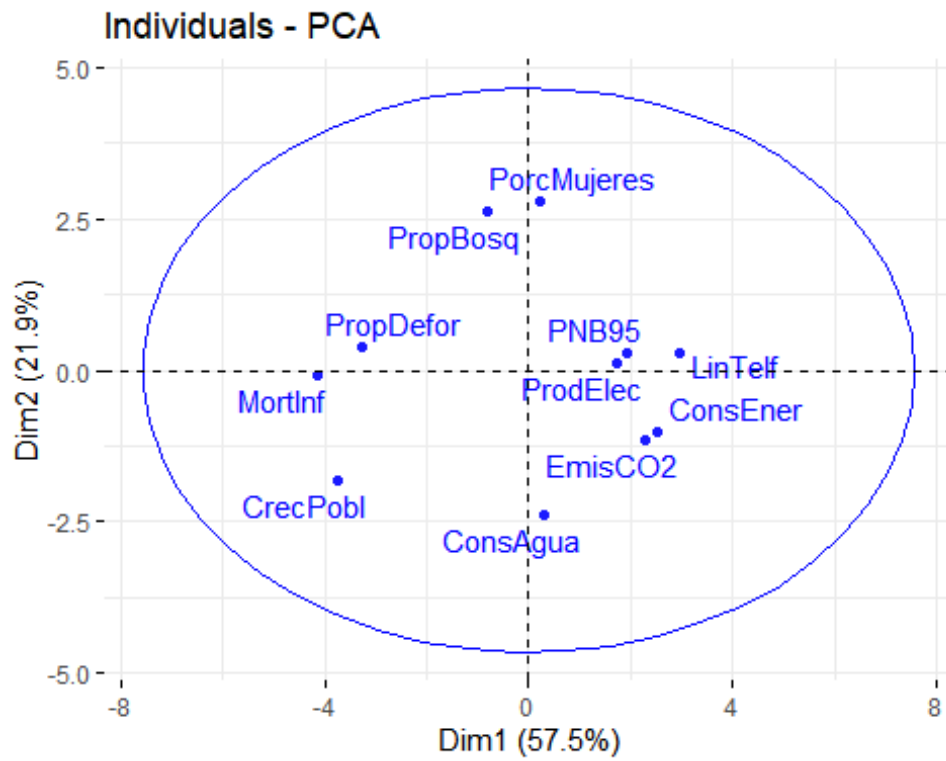
```
# Interpretación de resultados en término de agrupación de variables:
## Para el componente uno: PorcMujeres, propBosq, propDefor, mortInf;
```

queda la posibilidad de: PNB, prodElec, y crecPob. Por la descripción de las variables podemos llamar a este componente 'índice de riqueza'
Para el componente dos: consEner, emisCO2, consAgua y crecPob. Por la descripción de las variables podemos llamar a este componente 'índice de ruralidad'

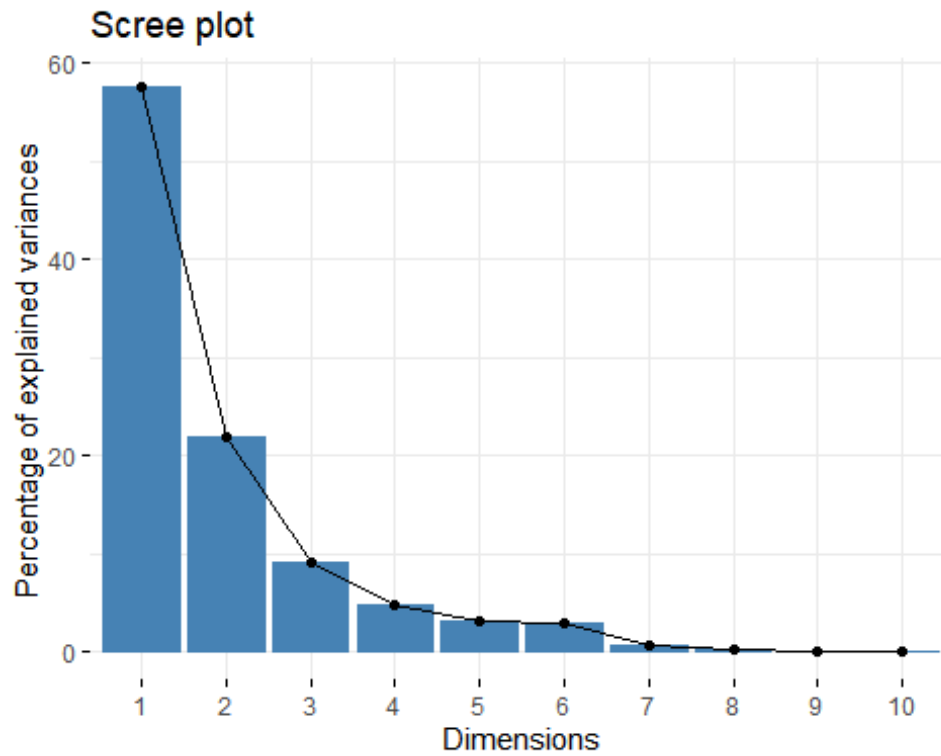
```
library(FactoMineR)
library(factoextra)
library(ggplot2)
datos=correlacion
cp3 = PCA(datos)
```



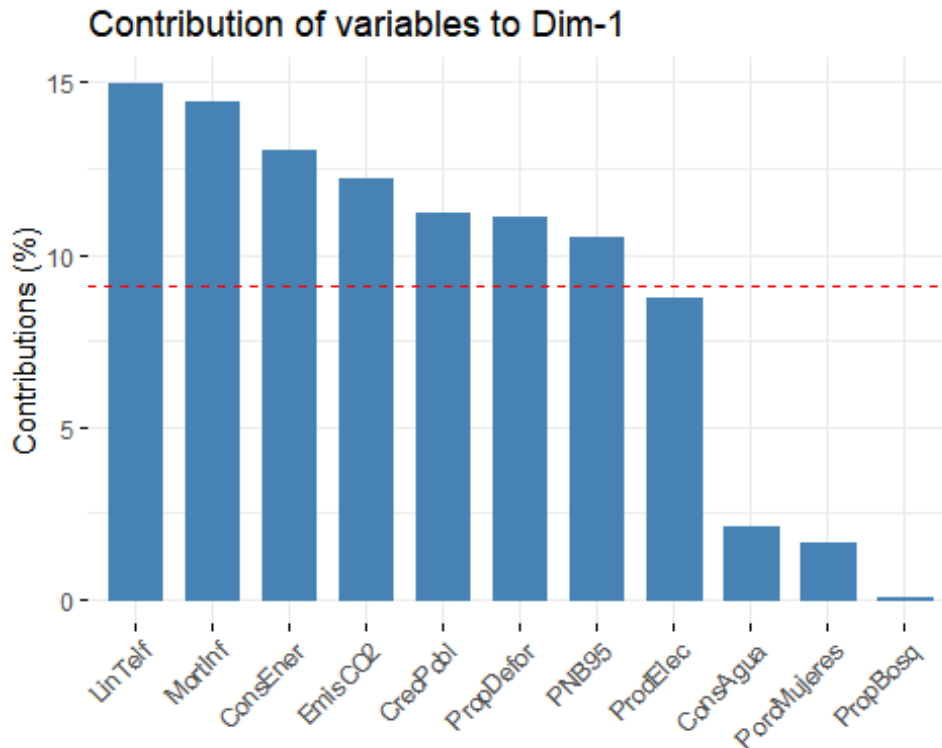
```
fviz_pca_ind(cp3, col.ind = "blue", addEllipses = TRUE, repel = TRUE)
```



```
fviz_screepLOT(cp3)
```



```
fviz_contrib(cp3, choice = c("var"))
```



Interpretación de gráficos:

En el primer gráfico, con el uso de la librería, se puede notar el porcentaje de proporción de varianza explicada que tiene cada uno de los dos primeros componentes (siendo mayormente exacto); se puede notar que para el primero es de 57.55% y para el segundo 21.92%. También, permite definir las variables que tienen una mayor influencia en cada una de ellas.

El segundo y tercer gráfico permite entender el mismo resultado descrito en el punto anterior, con una visualización distinta (agrupada por cuadrantes)

El penúltimo gráfico permite visualizar la proporción de varianza explicada en cada componente, donde se puede demostrar que los dos primeros son los que mayormente explican los datos.

El último gráfico permite visualizar el porcentaje de contribución (peso) de cada variable en el primero componente.